BMC Genomics

BioMed Central

Research article

The EH1 motif in metazoan transcription factors

Richard R Copley*

Address: Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK

Email: Richard R Copley* - copley@well.ox.ac.uk

* Corresponding author

Published: 27 November 2005

BMC Genomics 2005, **6**:169 doi:10.1186/1471-2164-6-169

Received: 26 August 2005

Accepted: 27 November 2005

This article is available from: http://www.biomedcentral.com/1471-2164/6/169

Abstract

Background: The Engrailed Homology 1 (EH1) motif is a small region, believed to have evolved convergently in homeobox and forkhead containing proteins, that interacts with the *Drosophila* protein groucho (*C. elegans* unc-37, Human Transducin-like Enhancers of Split). The small size of the motif makes its reliable identification by computational means difficult. I have systematically searched the predicted proteomes of *Drosophila*, *C. elegans* and human for further instances of the motif.

Results: Using motif identification methods and database searching techniques, I delimit which homeobox and forkhead domain containing proteins also have likely EH1 motifs. I show that despite low database search scores, there is a significant association of the motif with transcription factor function. I further show that likely EH1 motifs are found in combination with T-Box, Zinc Finger and Doublesex domains as well as discussing other plausible candidate associations. I identify strong candidate EH1 motifs in basal metazoan phyla.

Conclusion: Candidate EH1 motifs exist in combination with a variety of transcription factor domains, suggesting that these proteins have repressor functions. The distribution of the EH1 motif is suggestive of convergent evolution, although in many cases, the motif has been conserved throughout bilaterian orthologs. Groucho mediated repression was established prior to the evolution of bilateria.

Background

The Engrailed Homology 1 (EH1) motif is a short (<10 amino acids) region, initially found in engrailed (en) and other homeobox containing proteins, that mediates transcriptional repression via interaction with the WD40 repeat containing groucho (Gro) [1,2]. Shimeld [3] proposed that the EH1 motif of Smith and Jaynes was shared with various forkhead (FH/HNF-3) containing transcription factors. The short size of the motif, however, suggests that it may occur by chance in many different protein families. Shimeld did not demonstrate statistically significant sequence similarity between the motifs from the homeobox- and forkhead-containing families. However, the human orthologs of groucho (the transducin-like enhancer of split proteins) have been shown to interact with FOXA2 via a region of sequence containing an EH1 motif, clearly demonstrating the biological relevance of the sequence similarity [4].

In this article I search systematically for instances of the EH1 motif in homeobox and forkhead containing genes and go on to demonstrate that the EH1 motif is also found in proteins containing T-box, Doublesex Motif (DM) and Zn finger domains. I show that within metazoan

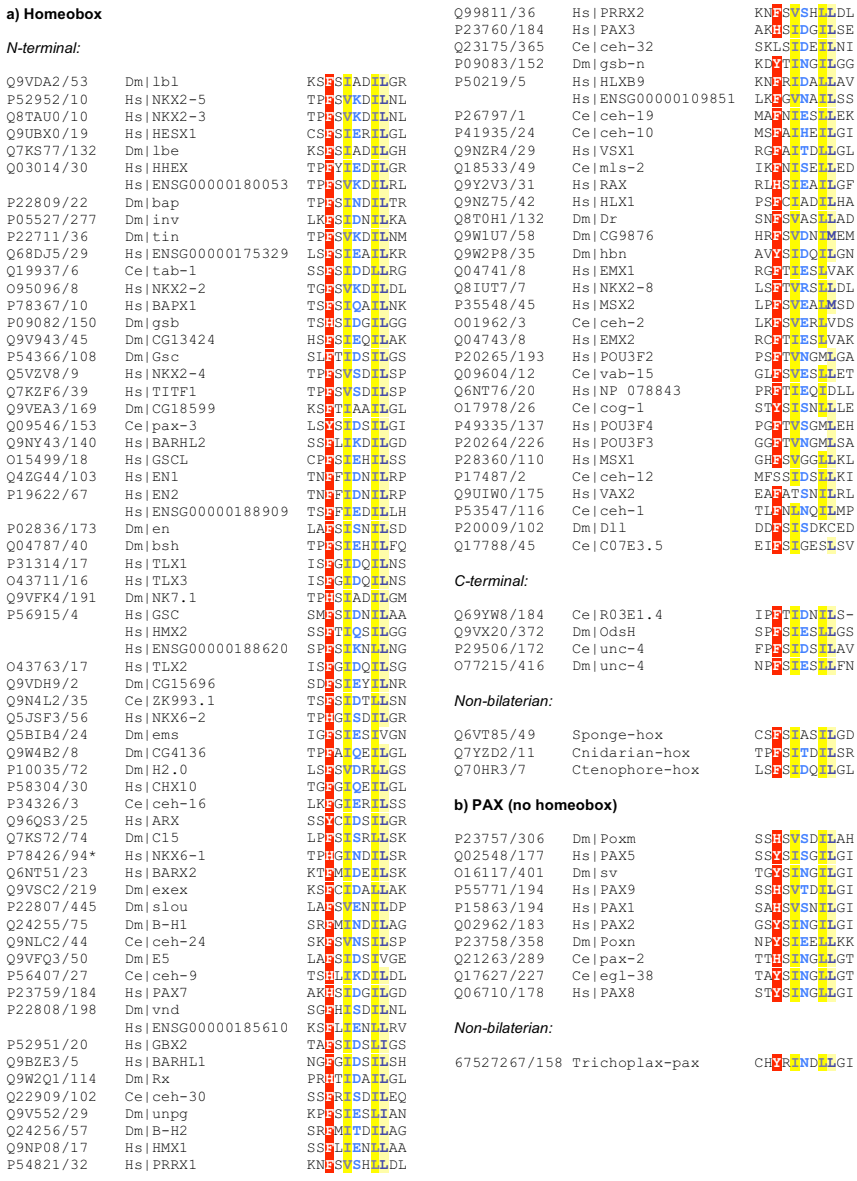

Figure 1

Alignments of putative EH1 motifs in a) Homeobox and b) Paired box domain containing proteins, subdivided by domain partners and orientation, with representative non-bilaterian sequences included. Alignments were derived from meme searches, as described in text. Conserved aromatic residues (FHYW) are coloured white on a red background ('a' in the consensus). Conserved aliphatic residues (ILV), black on a yellow background. ('l' in the consensus) Conserved big residues (EFHIKLMQRWY) blue on a light yellow background ('b' in the consensus). Conservation is calculated over the full alignment of sequences in figures 1 and 2. The figure was produced using the Chroma program [41]. Gene names are standard HUGO Gene Nommenclature Committee, flybase or wormbase symbols where available, otherwise accessions for their respective databases. When available Uniprot protein accessions are also given [42], along with the starting residue of the motif.

genomes, the observed association of the motif with transcription factor function is statistically significant. The location of the motif in members of the same transcription factor family is often non-homologous, occurring both N- and C-terminal to the DNA binding domain, suggesting that the presence of the motif is, in part, due to convergent evolution, as proposed by Shimeld; the conservation within orthologs points to many of these convergences predating the last common ancestor of the bilateria.

Results and Discussion

Significant association of EH1 motif with transcription factor function

I searched for sequence motifs in homeobox containing transcription factors taken from the proteins of human, *Drosophila melanogaster* and *Caenorhabditis elegans*, by first masking known Pfam domains [5], and then using the expectation maximization algorithm implemented in the meme program [6]. The first non-subfamily specific motif identified corresponded to previously known examples and new instances of, the EH1 motif (see Figure 1a), in 100 sites, with an E-value of $< 10^{-126}$. I then applied the same approach to Forkhead containing transcription factors, identifying 25 sites with a combined E-value of $< 10^{-31}$ (Figure 2a). These motifs also appeared to conform to the consensus of the EH1 motif, as previously reported by Shimeld [3].

To further investigate the significance of this similarity, I constructed hidden Markov models (HMM) of the motif ($EH1^{hox}$ & $EH1^{fh}$) which I then searched against the complete set of predicted proteins from human, *D. melanogaster* &*C. elegans*. The highest scoring non homeobox containing domain match of $EH1^{hox}$ was a Forkhead protein (human FOXL1), and the second highest scoring non-Forkhead containing match of $EH1^{fh}$ was to a homeobox containing protein (*D. melanogaster* inv). In both cases, nearly all the high scoring hits were to proteins containing domains with transcription factor function (see Figure 3). Among the best scoring matches of the $EH1^{hox}$ searches were several T-box (TBOX), Doublesex Motif (DM), Zinc finger (ZnF_C2H2) and ETS containing proteins (domain names as per SMART, Figure 2b–e) [7,8]. Excluding hits to homeobox containing proteins, but otherwise including all scores, the overall significance of the association of transcription factor function with higher scores to the $EH1^{hox}$ HMM is $P < 10^{-47}$, using a logistic regression model which tests association between score and transcription factor annotation (see methods and supplementary file 1 for raw data). The association remains significant when scores derived from Forkhead and PAX domain containing proteins are also excluded ($P < 10^{-34}$). This indicates that, although the scores associated with any individual EH1-like motif may not be statistically significant, overall, we would not see so many EH1-like sequences co-occurring with DNA binding domains if their co-occurrence were governed simply by chance – there is, therefore, likely to be a functional reason for these partnerships. In the following sections, I review the higher scoring associations detected here in the light of known gene functions.

EH1 motifs in homeobox and forkhead containing proteins

The presence of EH1 motifs within various homeobox, and to a lesser extent, forkhead containing proteins has been widely reported, although not systematically studied [3]. I found EH1-like motifs co-occurring with 3 major groupings of homeobox sub-types: the extended-hox class, typified by *Drosophila* engrailed (en); the paired class, including *Drosophila* goosecoid (gsc), and the NK class, including *Drosophila* tinman (tin) [1,9,10] (see [11] for a description of these broad classes). Related to the paired class homeobox domains, a number of genes containing PAIRED domains only (i.e. the PAX domain of SMART [7]) were also found to contain EH1-like motifs (see Figure 1b). With only a few exceptions, outlined below, the EH1-like motif occurs N-terminal to the homeobox domain and C-terminal to the PAIRED domain when present. A number of these proteins have been shown to interact with groucho or its orthologs e.g. *C. elegans* cog-1 [12], vertebrate Nkx proteins [13], *Drosophila* engrailed (en) and goosecoid (gsc) [2,14], and in high throughput assays *Drosophila* invected (inv) and and ladybird late (Ibl) [15].

A handful of EH1-like motifs are found C-terminal to homeobox domains. Of these, the best characterized is *C. elegans* unc-4, which has been shown to interact with the groucho ortholog unc-37 [16]; the *Drosophila* ortholog unc-4 also interacts with groucho in high throughput experiments [15]. The C-terminal EH1-like motif is conserved in the closely related *Drosophila* paralog OdsH. The gene prediction for the human ortholog of unc-4 (ensembl gene identifier ENSG00000164853) appears to be artefactually truncated, but the mouse ortholog (Uncx4.1 ENSMUSG00000029546) and corrected human gene models, contain EH1-like motifs both N & C-terminal to the homeobox domain. Taken together with the fact that in the majority of related homeobox containing proteins the EH1-like motifs are N-terminal, this suggests that the N-terminal motif has been lost in *Drosophila* and *C. elegans* unc-4 orthologs.

EH1-like motifs also occur N- and C-terminal to Forkhead domains. The N-terminal class consists of the sloppy-paired genes (slp1 and slp2) of *Drosophila* and orthologous or closely related sequences: human FOXG1, and *Drosophila* CG9571; the *C. elegans* ortholog fkh-2 contains an EH1-like motif although a cysteine residue causes a low

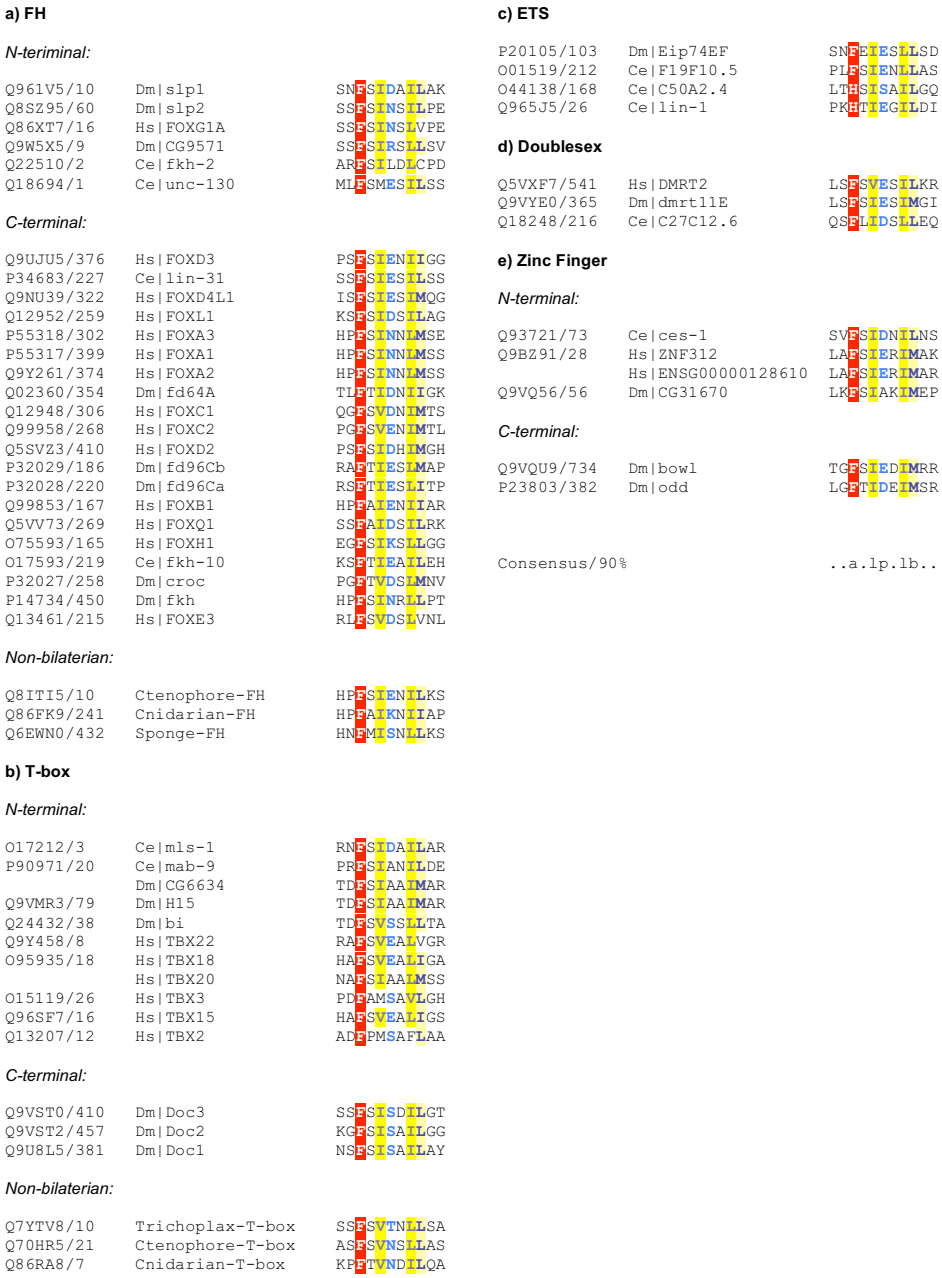

Figure 2
Alignments of putative EH1 motifs in a) Forkhead b) T-box c) ETS d) Doublesex and e) Zinc finger containing proteins. Alignment 'a' was derived from a meme search, as described in text. Sub-alignments b-e were derived from HMMER searches with the EH1^{hox} HMM. Other details as for Figure 1.

score. The C-terminal class consists of an apparent clade including the human FOXA, FOXB, FOXC and FOXD genes (Figure 2a), although if the EH1 motif was present in the common ancestor of this clade, multiple losses must have later occurred (see [17] for a Forkhead domain phylogeny). The situation is complicated somewhat by an EH1-like motif at the N-terminus of *C. elegans* unc-130 i.e. in the FOXD like family. The EH1 motif in slp1 has been shown to interact with groucho [18], and FOXA type genes have been shown to interact with human groucho orthologs [4].

EH1 motifs in novel domain contexts

Assuming a conservative per-domain cutoff score of 10.0 bits for true matches to the EH1^{hox} model (see Figure 3), yields hits to proteins containing T-box domains (highest score 13.1 bits); Doublesex (DM) domains (highest score 11.6 bits) and C2H2 Zinc fingers (highest score 11.2 bits). Also of note was a further match at 9.4 bits, to an ETS domain containing protein. Prompted by these similarities I further investigated the presence of EH1-like motifs in these families, looking for high scoring matches to the EH1^{hox} HMM that were conserved in closely related genes.

T-box containing proteins

I identified likely EH1 motifs co-occurring with T-Box domains in two distinct contexts (Figure 2b). The motif occurs C-terminal to the T-box in the *Drosophila* dorsocross proteins Doc1, Doc2 and Doc3. It is found N-terminal to the T-box in 11 proteins including mls-1 and mab-9 from *C. elegans;* H15, mid/nmr2 and bi/omd from *Drosophila*; in humans there are strong matches to TBX18, TBX20 and TBX22 and more marginal matches to TBX3 and TBX2. Although, to the best of my knowledge, none of these proteins has been shown to interact with groucho or its orthologs, several are known to act as transcriptional repressors: for instance, in murine heart development, Tbx20 represses Tbx2 which in turn represses Nmyc [19,20]; the Dorsocross genes from *Drosophila* repress wingless and ladybird [21], and Doc itself is repressed by mid/nmr2 [22]. The human proteins TBX1 and TBX10, and *Drosophila* org-1 which are closely related to those above, do not appear to contain EH1 motifs. The human T (brachyury) protein contains a motif broadly similar to the EH1 consensus: LQ**Y**R**V**DH**LL**SA in a comparable N-terminal location to those found in other T-box containing proteins. Although this motif scores poorly against EH1^{hox} (-0.1 bits), the homologous regions from other T orthologs (for instance, the non-bilaterian sequences discussed below) provide a more persuasive case for the presence of a functioning EH1 motif in these proteins.

Zinc finger containing proteins

The highest scoring match of EH1^{hox} to a C2H2 zinc finger containing protein, was ces-1 from *C. elegans* (bit score 11.2); this protein interacts with the groucho ortholog unc-37 [[23], #54] and can act as a repressor [24]. The putative EH1 motif is at the N-terminal end of ces-1. In contrast, the *Drosophila* proteins bowl and odd have EH1-like motifs at their C-terminal ends (with bit scores of 10.9 & 8.4 respectively). In neither case is there direct evidence from high throughput studies of an interaction with groucho, but both can function as repressors [25]. The human protein ZNF312 (bit score 8.6) is the ortholog of zebrafish Fezl, which contains an EH1 motif essential for repressor activity [26] – this motif is conserved in the human paralog ENSG00000128610 and likely *Drosophila* ortholog CG31670 (bit scores of 8.4 & 5.1) (Figure 2e).

Doublesex motif containing proteins

The Doublesex Motif (DM) was first found in proteins controlling sexual differentiation in *Drosophila*. Two DM containing proteins were confidently predicted to contain EH1-like motifs – human DMRT2 (bit score 11.6), and *Drosophila* dmrt11e (bit score 11.2) – these are likely orthologs; a *C. elegans* protein, C27C12.6 contained a weaker match (bit score 6.6) (Figure 2d). The molecular function of these proteins is unknown.

Other potential associations with transcription factor domains

Although scoring less highly than some non-transcription factor hits, another intriguing association is with the ETS domain. The three uncharatcerized *C. elegans* paralogs F19F10.5, F19F10.1 & C50A2.4 contain C-terminal matches to the EH1 motifs (bit scores 9.4, 2.3 & 7.4), and two other ETS proteins, *C. elegans* lin-1, and *Drosophila* Eip74EF, both have relatively high scoring matches (bit scores 6.5 & 6.6) (Figure 2c). A high scoring protein that is not annotated as a transcription factor (as it contains no interpro domains) is *Drosophila* Hairless (H) with a score of 8.3 bits. Experimental work has previously confirmed the presence of an EH1-like motif (SS**Y**S**I**HS**LL**GG) within H that is responsible for its interaction with groucho [27]. The *Drosophila* protein Dorsal has been reported to interact with groucho via an EH1-like motif [28] – this region (NG**P**T**L**SN**LL**SF) is markedly different to those reported here, having a low score against EH1^{hox} (-10.7 bits) and so may better be regarded as a, so far, unique type of groucho interaction motif.

Evolutionary considerations

Convergent evolution

The EH1 motif is found N- and C-terminal to homeobox, forkhead, T-box and Zn finger protein domains. Clearly, as the locations of the EH1 motif are non-homologous, the N- and C-terminal associations must have occurred independently. The short size of the motif makes it tempting to speculate that the motif itself may have arisen independently (i.e. in repeated cases it may have evolved within sequence that was already part of the gene, rather

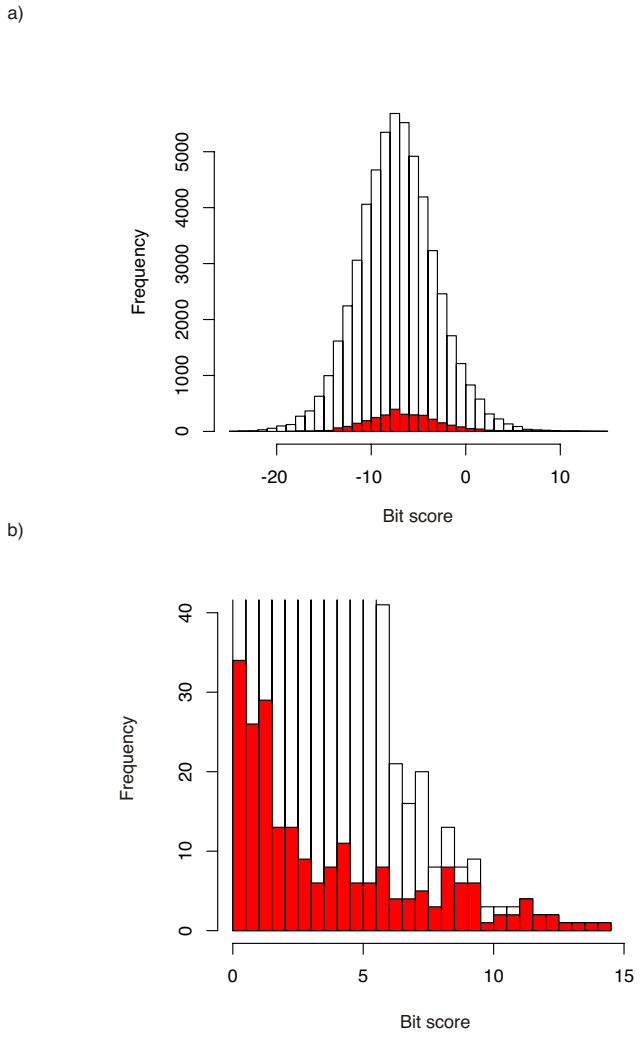

Figure 3
a) Distribution of HMMER bit scores for the database search of EH1^{hox} HMM against the combined proteomes of human, *D. melanogaster* and *C. elegans*. Counts from scores from transcription factors (see methods) have been coloured red – i.e. the proportion of a bar coloured red is equal to the proportion of transcription factors. Scores from proteins containing a homeobox domain (interpro accession IPR001356), from which the EH1^{hox} HMM was derived, have been excluded, b) as for 'a', but rescaled to show region of biological relevance. High scoring hits are greatly enriched in specific transcription factor families. For scores ≥ 5.0 bits, there are 68 transcription factors and 142 non-transcription factors; for scores <5.0 bits, 3075 transcription factors and 51513 non-transcription factors giving a chi-square test p-value statistic of $P < 0.0001$ – the statistical significance is discussed more fully in the text.

than via a recombination event). The strongest evidence for this is that, in general, the majority of domain combinations occur in a fixed N to C orientation, suggesting that recombination events combining domains are relatively rare [29,30]. The fact that we would here have many such events suggests that the alternative hypothesis of independent invention is more appropriate.

Pre-bilaterian origins of association with different transcription factors

Groucho is orthologous to the *C. elegans* unc-37 gene, and the four human paralogs TLE1-4 (Transducin Like Enhancer of split). An ortholog is also found in the cnidarian *Hydra mangipapillata* (e.g. the EST with gi 47137860, data not shown), and certain cnidarian homeobox containing genes also contain an EH1-like motif, suggesting groucho/EH1 mediated repression pre-dates the split between diplobasts and triplobasts; indeed, a sponge Bar/Bsh like homeobox containing protein (i.e. protein gi: 33641772) [31] also contains an EH1-like motif, as does paxb from the non-bilaterian placozoan *Trichoplax adhaerens* [32] and a Tlx-like protein from a ctenophore (gi: 38602653), suggesting the repression system was in place in the earliest animals (see [33] for a discussion of early metazoan evolution). I find high scoring EH1-like motifs in Forkhead domain containing proteins from sponges, cnidarians and ctenophores, in both the C-terminal (FOXA-D clade) (region II in [34]) and N-terminal (FOXG, sloppy paired clade) varieties (reported as 'HPFSI' in [35]). The presumed ortholog of 'T' from the *Trichoplax adhaerens* [36] contains an EH1-like motif (8.6 bits). These results suggest that groucho mediated repression using a variety of transcription factors was widespread in the last common ancestor of the metazoa.

Conclusion

Candidate EH1 motifs exist in combination with a variety of transcription factor domains, suggesting that these proteins have roles as repressors of transcriptional activity. The distribution of the EH1 motif is suggestive of a number of instances of convergent evolution, although in many cases the motif has been conserved throughout bilaterian orthologs. Together with the existence of a cnidarian Groucho ortholog, this leads to the conclusion that EH1/Groucho mediated repression was established prior to the evolution of bilateria.

Methods

Proteomes were derived from ensembl 32 (human NCBI 35, *C. elegans* wormbase 140, *Drosophila* BDGP 4) [37]. In cases of multiple splice variants, the one with the most exons was included (or the longest in the case of ties). Transcription factor activity was taken as the presence of the gene ontology accession GO:0003700 associated with an interpro domain predicted for the protein [38]. These data were also taken from ensembl. Although C2H2 subtype Zn fingers are not annotated by Interpro as transcription factors they are DNA binding and frequently have

this role, so have been included in the transcription factor set. Bit scores reported in the text are for comparisons of the EH1^{hox} HMM against the target sequence using the HMMER software package [39].

The association of transcription factor function (coded as a dichotomous variable, *t*, taking the values 1 [transcription factor] or 0 [non-transcription factor]) with the bit score, *x*, of the EH1^{hox} HMM, was tested using a logistic regression model implemented in the glm() function of the R package [40]). I fitted the model

$$\text{Prob}(t = 1) = \exp(a + bx)/(1 + \exp(a + bx))$$

The coefficients *a*, *b* were estimated from the data by maximum-likelihood. The hypothesis of no association is equivalent to testing if $b = 0$.

Where inferences of orthology are made, they are based on clear-cut separation of BLAST scores or alignment-based phylogenies.

Authors' contributions

RRC performed the analysis and wrote the paper.

Additional material

Additional File 1

Each non-homeobox containing protein in the database is classified as either being a transcription factor or not (see methods), and its score against the EH1^{hox} HMM is given.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-6-169-S1.txt]

Acknowledgements

I am grateful to an anonymous referee for comments on TBX15 & brachyury. I thank the Wellcome Trust for financial support, Dr. Richard Mott for statistical advice, Drs. Martin Taylor and William Valdar for helpful suggestions.

References

1. Smith ST, Jaynes JB: **A conserved region of engrailed, shared among all en-, gsc-, Nk1-, Nk2- and msh-class homeoproteins, mediates active transcriptional repression in vivo.** *Development* 1996, **122(10):**3141-3150.
2. Tolkunova EN, Fujioka M, Kobayashi M, Deka D, Jaynes JB: **Two distinct types of repression domain in engrailed: one interacts with the groucho corepressor and is preferentially active on integrated target genes.** *Mol Cell Biol* 1998, **18(5):**2804-2814.
3. Shimeld SM: **A transcriptional modification motif encoded by homeobox and fork head genes.** *FEBS Lett* 1997, **410(2–3):**124-125.
4. Wang JC, Waltner-Law M, Yamada K, Osawa H, Stifani S, Granner DK: **Transducin-like enhancer of split proteins, the human homologs of Drosophila groucho, interact with hepatic nuclear factor 3beta.** *J Biol Chem* 2000, **275(24):**18418-18423.
5. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, *et al.*: **The Pfam protein families database.** *Nucleic Acids Res* 2004:D138-141.
6. Bailey TL, Elkan C: **Fitting a mixture model by expectation maximization to discover motifs in biopolymers.** *Proc Int Conf Intell Syst Mol Biol* 1994, **2:**28-36.
7. **SMART – Simple Modular Architecture Research Tool** [http://smart.embl.de]
8. Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P: **SMART 4.0: towards genomic data integration.** *Nucleic Acids Res* 2004:D142-144.
9. Galliot B, de Vargas C, Miller D: **Evolution of homeobox genes: Q50 Paired-like genes founded the Paired class.** *Dev Genes Evol* 1999, **209(3):**186-197.
10. Jagla K, Bellard M, Frasch M: **A cluster of Drosophila homeobox genes involved in mesoderm differentiation programs.** *Bioessays* 2001, **23(2):**125-133.
11. Banerjee-Basu S, Baxevanis AD: **Molecular evolution of the homeodomain family of transcription factors.** *Nucleic Acids Res* 2001, **29(15):**3258-3269.
12. Chang S, Johnston RJ Jr, Hobert O: **A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of C. elegans.** *Genes Dev* 2003, **17(17):**2123-2137.
13. Muhr J, Andersson E, Persson M, Jessell TM, Ericson J: **Groucho-mediated transcriptional repression establishes progenitor cell pattern and neuronal fate in the ventral neural tube.** *Cell* 2001, **104(6):**861-873.
14. Jimenez G, Verrijzer CP, Ish-Horowicz D: **A conserved motif in goosecoid mediates groucho-dependent repression in Drosophila embryos.** *Mol Cell Biol* 1999, **19(3):**2080-2087.
15. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, *et al.*: **A protein interaction map of Drosophila melanogaster.** *Science* 2003, **302(5651):**1727-1736.
16. Winnier AR, Meir JY, Ross JM, Tavernarakis N, Driscoll M, Ishihara T, Katsura I, Miller DM 3rd: **UNC-4/UNC-37-dependent repression of motor neuron-specific genes controls synaptic choice in Caenorhabditis elegans.** *Genes Dev* 1999, **13(21):**2774-2786.
17. Mazet F, Yu JK, Liberles DA, Holland LZ, Shimeld SM: **Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria.** *Gene* 2003, **316:**79-89.
18. Andrioli LP, Oberstein AL, Corado MS, Yu D, Small S: **Groucho-dependent repression by sloppy-paired 1 differentially positions anterior pair-rule stripes in the Drosophila embryo.** *Dev Biol* 2004, **276(2):**541-551.
19. Stennard FA, Costa MW, Lai D, Biben C, Furtado MB, Solloway MJ, McCulley DJ, Leimena C, Preis JI, Dunwoodie SL, *et al.*: **Murine T-box transcription factor Tbx20 acts as a repressor during heart development, and is essential for adult heart integrity, function and adaptation.** *Development* 2005, **132(10):**2451-2462.
20. Cai CL, Zhou W, Yang L, Bu L, Qyang Y, Zhang X, Li X, Rosenfeld MG, Chen J, Evans S: **T-box genes coordinate regional rates of proliferation and regional specification during cardiogenesis.** *Development* 2005, **132(10):**2475-2487.
21. Reim I, Lee HH, Frasch M: **The T-box-encoding Dorsocross genes function in amnioserosa development and the patterning of the dorsolateral germ band downstream of Dpp.** *Development* 2003, **130(14):**3187-3204.
22. Reim I, Mohler JP, Frasch M: **Tbx20-related genes, mid and H15, are required for tinman expression, proper patterning, and normal differentiation of cardioblasts in Drosophila.** *Mech Dev* 2005.
23. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, *et al.*: **A map of the interactome network of the metazoan C. elegans.** *Science* 2004, **303(5657):**540-543.
24. Thellmann M, Hatzold J, Conradt B: **The Snail-like CES-1 protein of C. elegans can block the expression of the BH3-only cell-death activator gene egl-1 by antagonizing the function of bHLH proteins.** *Development* 2003, **130(17):**4057-4071.
25. Campbell G: **Regulation of gene expression in the distal region of the Drosophila leg by the Hox11 homolog, C15.** *Dev Biol* 2005, **278(2):**607-618.
26. Levkowitz G, Zeller J, Sirotkin HI, French D, Schilbach S, Hashimoto H, Hibi M, Talbot WS, Rosenthal A: **Zinc finger protein too few controls the development of monoaminergic neurons.** *Nat Neurosci* 2003, **6(1):**28-33.

27. Barolo S, Stone T, Bang AG, Posakony JW: **Default repression and Notch signaling: Hairless acts as an adaptor to recruit the corepressors Groucho and dCtBP to Suppressor of Hairless.** *Genes Dev* 2002, **16(15):**1964-1976.
28. Flores-Saaib RD, Jia S, Courey AJ: **Activation and repression by the C-terminal domain of Dorsal.** *Development* 2001, **128(10):**1869-1879.
29. Apic G, Gough J, Teichmann SA: **Domain combinations in archaeal, eubacterial and eukaryotic proteomes.** *J Mol Biol* 2001, **310(2):**311-325.
30. Gough J: **Convergent evolution of domain architectures (is rare).** *Bioinformatics* 2005, **21(8):**1464-1471.
31. Hill A, Tetrault J, Hill M: **Isolation and expression analysis of a poriferan Antp-class Bar-/Bsh-like homeobox gene.** *Dev Genes Evol* 2004, **214(10):**515-523.
32. Hadrys T, Desalle R, Sagasser S, Fischer N, Schierwater B: **The Trichoplax PaxB Gene: A Putative Proto-PaxA/B/C Gene Predating the Origin of Nerve and Sensory Cells.** *Mol Biol Evol* 2005, **22(7):**1569-1578.
33. Medina M, Collins AG, Silberman JD, Sogin ML: **Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA.** *Proc Natl Acad Sci USA* 2001, **98(17):**9707-9712.
34. Adell T, Muller WE: **Isolation and characterization of five Fox (Forkhead) genes from the sponge Suberites domuncula.** *Gene* 2004, **334:**35-46.
35. Yamada A, Martindale MQ: **Expression of the ctenophore Brain Factor 1 forkhead gene ortholog (ctenoBF-1) mRNA is restricted to the presumptive mouth and feeding apparatus: implications for axial organization in the Metazoa.** *Dev Genes Evol* 2002, **212(7):**338-348.
36. Martinelli C, Spring J: **Distinct expression patterns of the two T-box homologues Brachyury and Tbx2/3 in the placozoan Trichoplax adhaerens.** *Dev Genes Evol* 2003, **213(10):**492-499.
37. Hubbard T, Andrews D, Caccamo M, Cameron G, Chen Y, Clamp M, Clarke L, Coates G, Cox T, Cunningham F, *et al.*: **Ensembl 2005.** *Nucleic Acids Res* 2005:D447-453.
38. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bradley P, Bork P, Bucher P, Cerutti L, *et al.*: **InterPro, progress and status in 2005.** *Nucleic Acids Res* 2005:D201-205.
39. **HMMER: sequence analysis using profile hidden Markov models** [http://hmmer.wustl.edu]
40. **The R project for statistcal computing** [http://www.r-project.org]
41. Goodstadt L, Ponting CP: **CHROMA: consensus-based colouring of multiple alignments for publication.** *Bioinformatics* 2001, **17(9):**845-846.
42. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, *et al.*: **The Universal Protein Resource (UniProt).** *Nucleic Acids Res* 2005:D154-159.